J Am Med Inform Assoc 2021:(submitted)
draft material -- please do not cite

Feasibility of Replacing ICD-10-CM with ICD-11 – A Content Analysis

Kin Wah Fung[1], MD, MS, MA; Julia Xu[1], MD, PhD; Shannon McConnell-Lamptey[2], MS, RHIA; Donna Pickett[2], MPH, RHIA; Olivier Bodenreider[1], MD, PhD

Affiliations: [1]*National Library of Medicine, National Institutes of Health, USA;* [2]*National Center for Health Statistics, Centers for Disease Control and Prevention, USA*

Correspondence and reprints:

Kin Wah Fung
Building 38A, Rm9S918, MSC-3826
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894

Telephone: 301 – 827 5001
Fax: 301 – 496 0663
Email: kwfung@nlm.nih.gov

Keywords: ICD-11, ICD-10, ICD-10-CM, Controlled medical vocabularies, Medical terminologies

Word count: 4,141

ABSTRACT

Objective: To assess the feasibility of replacing ICD-10-CM with ICD-11 for morbidity coding based on content analysis and coverage

Materials and Methods: Most frequently used ICD-10-CM codes from each chapter covering 60% of usage were identified from Medicare claims and hospital data. Each ICD-10-CM code was recoded in ICD-11, using postcoordination (combination of codes) if necessary. Recoding was done by two terminologists independently and differences discussed to reach consensus. Failure analysis was done for cases where full representation was not achieved even with postcoordination. After recoding, the coding guidance (inclusions, exclusions and index) of the ICD-10-CM and ICD-11 codes were reviewed for conflict.

Results: Overall, 23.5% of 943 codes could be fully represented by ICD-11 without postcoordination, and a further 8.6% could be fully postcoordinated. With the addition of nine extension codes, it is possible to increase the proportion that can be fully postcoordinated from 8.6% to 35.2%. Coding guidance review identified potential conflicts in 10% of codes, but mostly not affecting recoding.

Conclusion: With some minor enhancements to postcoordination, ICD-11 can fully represent almost 60% of the most frequently used ICD-10-CM codes. This coverage is far superior to the 24.3% exact representation of ICD-9-CM codes by ICD-10-CM. Therefore, ICD-11 should be considered as a candidate to replace ICD-10-CM for morbidity coding.

INTRODUCTION

The International Classification of Diseases (ICD) has been in use for collection of global health trends and statistics for over a century. [1, 2] Its latest version, ICD-11, was adopted in May 2019 and will be implemented in member countries of the World Health Organization (WHO) starting in January 2022. [3-5] Due to specific requirements in some countries, over two dozen national extensions of ICD have been developed for past versions of ICD. In the US, the first version of the national extension known as Clinical Modification (CM) was ICD-9-CM released in 1979. According to the official documentation of ICD-9-CM, “the term "clinical" is used to emphasize the modification's intent: to serve as a useful tool in the area of classification of morbidity data for indexing of medical records, medical care review, and ambulatory and other medical care programs, as well as for basic health statistics. To describe the clinical picture of the patient, the codes must be more precise than those needed only for statistical groupings and trend analysis.” [6] The same practice of modifying the international ICD core for clinical purpose continued in ICD-10-CM, which replaced ICD-9-CM in 2015.

The main advantage of developing a US national extension is the ability to add necessary detail under the framework of the international core to serve clinical and administrative (e.g., reimbursement) needs. Another advantage is that updates to the national extension can happen more frequently, as ICD-10-CM is updated yearly compared to the three-year cycle for ICD-10. However, there are potential drawbacks. Firstly, significant effort is involved in maintaining an extension. Secondly, there is usually a delay between the release of the international version and the national extension. Moreover, there can be incongruence between the national extension and the international core. In principle, everything in the Clinical Modification should be totally

compatible with the parent system. However, some significant differences can be observed between ICD-10-CM and ICD-10. For example, the ICD-10 category *E14 Unspecified diabetes mellitus* is not present in ICD-10-CM, because diabetes mellitus of unspecified type is coded under *E11 Type 2 diabetes mellitus* by default. Another example is the addition to ICD-10-CM of a new category *K68 Disorders of retroperitoneum* that is not present in ICD-10.

Decades of research in controlled medical vocabularies and knowledge representation have resulted in better understanding of the principles and best practices in medical terminology management. [7] Some of these principles have been embraced by ICD-11. Apart from the introduction of the foundation component, the most noticeable novel feature in ICD-11 is postcoordination. [8, 9] Postcoordination is the combination of codes to represent new meaning - a powerful and efficient way to expand the coverage, expressivity and granularity of a terminology. Towards this end, ICD-11 offers 14,500 extension codes for postcoordination. This new capability, together with the considerable increase in the number of codes - 4,015 (37.9%) more codes than ICD-10, may lead one to question whether it is still necessary to develop a Clinical Modification for ICD-11. [10] In fact, the recommendations from the National Committee on Vital and Health Statistics (NCVHS) to the Secretary of the Department of Health and Human Services include research to determine whether ICD-11 can fully support morbidity classification in the US without development of a US clinical modification. [11]

In a previous study, we examined a limited sample of ICD-10-CM codes to see how well they could be represented in ICD-11. [10] The objective of the present investigation is to assess the feasibility of replacing ICD-10-CM with ICD-11 based on content analysis and coverage. More

specifically, we attempted to use ICD-11 to recode a representative sample of frequently used ICD-10-CM codes identified from Medicare claims and hospital data. In addition to determining whether the full meaning of an ICD-10-CM code could be represented in ICD-11, with or without postcoordination, we also reviewed the accompanying coding guidance (inclusions, exclusions and index) to identify subtle differences in code meaning that might not be conveyed by the code names and hierarchies alone. We believe this is the first extensive study on the feasibility of replacing ICD-10-CM with ICD-11.

MATERIALS AND METHODS

We first identified the most frequently used ICD-10-CM codes from insurance claims and hospital data, and then recoded them using ICD-11, using postcoordination if necessary. We assessed the coverage of ICD-11 and noted reasons for not achieving full representation. We then looked for potential conflicts caused by the accompanying coding guidance.

1. Most frequently used ICD-10-CM codes

We accessed Medicare claims data through the Virtual Research Data Center (VRDC) of the Centers for Medicare & Medicaid Services (CMS). [12] We gathered all ICD-10-CM codes in in-patient and out-patient claims, either as principal or secondary diagnoses, for the year 2017. Since there was a two-year lag before data was available through VRDC, 2017 was the latest full year of data at the time of the study. The data covered over 60 million Medicare beneficiaries, but since most Medicare patients were over 65, obstetric and pediatric codes from three chapters (chapter 15, 16 and 17) were missing. For these three chapters, we used data from three community hospitals (one tertiary, one secondary and one pediatrics tertiary care center)

affiliated with the University of Nebraska Medical Center. For this smaller data source, we used all data from October 2015 (beginning of ICD-10-CM use) to March 2020 to ensure maximal coverage. For each chapter, we identified the most frequently used ICD-10-CM codes that together covered 60% of unique patients. We excluded codes that were not valid in the 2021 version of ICD-10-CM, used as our reference for recoding to ICD-11. This study was rated as not human subject research by the Office of Human Research Protection at the National Institutes of Health.

2. Recoding ICD-10-CM codes in ICD-11

a. Best matching ICD-11 code(s), postcoordination and match type

For each ICD-10-CM code, we identified the best matching ICD-11 code valid for coding – the lowest level code. Recoding was done independently by JX and SM who are very knowledgeable in ICD-10-CM and ICD-11. All discrepancies were duly recorded and discussed until consensus was reached. We used the online ICD-11 browser to search the index and check the tabular listing. [13] Our recoding guidelines can be summarized as follows.

- Follow the ICD-11 coding reference guide for morbidity coding. [9]
- Ignore the parts of the ICD-10-CM or ICD-11 name that conveyed absence of information e.g., *gout* ***unspecified***, *Zoster* ***without complications***.
- Use ICD-11 codes that are equivalent or broader in meaning than the ICD-10-CM code (see Results-failure analysis for exceptions).
- For ICD-10-CM codes recoded to broader ICD-11 codes, attempt postcoordination, as allowed by the browser, to improve the match. For example, ICD-10-CM code *H52.13*

Myopia, bilateral was recoded as the broader ICD-11 code *9D00.0 Myopia*, but could be fully matched with postcoordination by adding the extension code *XK9J Bilateral*.

We then determined the degree of representation of the ICD-10-CM code by the ICD-11 recoding. There were three levels of representation: 1) full representation without postcoordination, 2) full representation with postcoordination, and 3) partial representation (postcoordination was not allowed or not sufficient to achieve full representation). This determination was based on the name and location in the organizational structure of the ICD-10-CM and ICD-11 codes. Of note, we did not consider matches in index or inclusion terms as indicative of full representation, as these terms were often narrower than the codes themselves. For example, the ICD-10-CM code *H54.8 Legal blindness, as defined in USA* was recoded as ICD-11 code *9D90.3 Severe vision impairment* that had "Legal blindness – USA" as an inclusion, but we still considered this partial representation because severe vision impairment remains broader than legal blindness defined in a specific country.

b. Failure analysis

All codes with partial representation were reviewed to determine the reason for failure and the type of missing information.

c. Coding guidance review

Both ICD-10-CM and ICD-11 use inclusion and exclusion terms to provide guidance to coders. Both also provide a comprehensive index to associate additional terms with the codes. The inclusion, exclusion and index terms are often more specific than the codes they are associated

with. In addition, most ICD-11 codes also have a textual description to define the meaning of the code. Apparently equivalent codes sometimes have different meanings because of the difference in the accompanying coding guidance.

i. Definitions, inclusions and exclusions

For all ICD-10-CM and ICD-11 codes under investigation, we reviewed the definition, inclusions and exclusions of the codes and their ancestors for possible conflicts. These conflicts could be between:

- ICD-11 textual description and inclusions or exclusions of the ICD-10-CM code and its ancestors
- inclusions of the ICD-10-CM code and its ancestors, and exclusions of the ICD-11 code and its ancestors (e.g., an inclusion in ICD-10-CM was an exclusion in ICD-11)
- exclusions of the ICD-10-CM code and its ancestors, and inclusions of the ICD-11 code and its ancestors

ii. Index terms

An index conflict occurred when an index term of the ICD-10-CM code was also found in the ICD-11 index, but pointing to a code other than the one selected in recoding. Since one ICD code could be associated with many index terms, it would be impractical to review all index terms. Therefore, we used normalized matching to reduce the workload of manual review. We extracted all index terms pointing to the ICD-10-CM codes, normalized them with the UMLS Lexical Tool, *luinorm*, [14, 15] then matched them to the ICD-11 normalized index terms. All cases in which the matched ICD-11 index term pointed to a code different from the one selected in recoding were reviewed.

RESULTS

1. Most frequently used ICD-10-CM codes

In the CMS data, there were 61 million unique patients and 28,981 unique ICD-10-CM codes. In the hospital data, there were 778,000 unique patients and 23,832 unique ICD-10-CM codes. Based on both data sources, we identified altogether 962 unique ICD-10-CM codes required to cover 60% of patients, of which 943 were still active in 2021. As shown in Table 1, the number of codes contributed by each chapter varied considerably, ranging from 5 (chapter 3) to 363 codes (chapter 19). This was determined by the size of the chapter (total number of codes) and the spread of usage. The last column in Table 1 is the percentage of codes required for 60% usage and is a measure of usage spread, which also shows significant variation among chapters. The full list of codes is available as online supplementary material (Appendix A).

Table 1. Distribution of most frequently used ICD-10-CM codes

Chapter	Code range	Total no. of codes	Top codes covering 60% usage	%
1	A00-B99 Certain infectious and parasitic diseases	1058	19	1.8%
2	C00-D49 Neoplasms	1661	66	4.0%
3	D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	*251*	*5*	2.0%
4	E00-E89 Endocrine, nutritional and metabolic diseases	908	10	1.1%
5	F01-F99 Mental, Behavioral and Neurodevelopmental disorders	747	10	1.3%
6	G00-G99 Diseases of the nervous system	622	13	2.1%
7	H00-H59 Diseases of the eye and adnexa	2606	51	2.0%
8	H60-H95 Diseases of the ear and mastoid process	653	18	2.8%
9	I00-I99 Diseases of the circulatory system	1378	14	1.0%
10	J00-J99 Diseases of the respiratory system	341	12	3.5%

11	K00-K95 Diseases of the digestive system	799	25	3.1%
12	L00-L99 Diseases of the skin and subcutaneous tissue	871	61	7.0%
13	M00-M99 Diseases of the musculoskeletal system and connective tissue	6487	43	0.7%
14	N00-N99 Diseases of the genitourinary system	672	10	1.5%
15	O00-O9A Pregnancy, childbirth and the puerperium	2267	45	2.0%
16	P00-P96 Certain conditions originating in the perinatal period	443	12	2.7%
17	Q00-Q99 Congenital malformations, deformations and chromosomal abnormalities	838	53	6.3%
18	R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	722	56	**7.8%**
19	S00-T88 Injury, poisoning and certain other consequences of external causes	**40654**	**363**	0.9%
20	V00-Y99 External causes of morbidity	6940	20	*0.3%*
21	Z00-Z99 Factors influencing health status and contact with health services	1266	37	2.9%
Total		72184	943	1.3%

Bold type – highest, *italics* – lowest

2. Recoding ICD-10-CM codes in ICD-11

a. Best matching ICD-11 code(s), postcoordination and match type

Overall, of the 943 codes, 222 (23.5%) could be fully represented without postcoordination, 81 codes (8.6%) could be fully represented with postcoordination, and the remaining 640 codes (67.9%) could only achieve partial representation. Due to the considerable difference in the number of codes among chapters, we also performed a chapter-based analysis. (Table 2) All codes from chapter 3 (blood and immune system) could be fully represented while none of the codes from chapter 19 (injury and poisoning) and 20 (external causes of morbidity) could. Across all chapters, an average of 47.1% of codes per chapter could be fully represented without postcoordination, corresponding to an average usage of 53.1%. Note that the usage percentage

here refers to the proportion of usage among the 943 ICD-10-CM codes in this study. Moreover, usage data cannot be aggregated across chapters since they come from two disparate data sources.

Agreement between the two terminologists in ICD-11 coding is summarized in Figure 1. Before discussion, agreement on the choice of ICD-11 main codes was observed in 716 (75.9%) cases. Among these 716 cases, postcoordination was used by both terminologists in 253 cases, in which they used the same postcoordination codes in 199 cases (78.7% agreement).

Table 2. Recoding ICD-10-CM codes in ICD-11, results by chapter

Chapter	**Full representation without postcoordination**		**Full representation with postcoordination**		**Partial representation**	
	% of codes	% of usage	% of codes	% of usage	% of codes	% of usage
1	**52.6%**	**70.1%**	21.1%	14.3%	26.3%	15.6%
2	**37.9%**	**46.8%**	36.4%	28.9%	25.8%	24.2%
3	**100.0%**	**100.0%**	0.0%	0.0%	0.0%	0.0%
4	**80.0%**	**86.6%**	10.0%	4.7%	10.0%	8.7%
5	**60.0%**	**54.5%**	0.0%	0.0%	40.0%	45.5%
6	**61.5%**	**55.8%**	0.0%	0.0%	38.5%	44.2%
7	17.6%	29.6%	13.7%	11.2%	**68.6%**	**59.2%**
8	16.7%	30.0%	**44.4%**	**37.4%**	38.9%	32.6%
9	**64.3%**	**87.6%**	7.1%	3.4%	28.6%	9.0%
10	**83.3%**	**92.8%**	0.0%	0.0%	16.7%	7.2%
11	**64.0%**	**81.2%**	8.0%	3.1%	28.0%	15.8%
12	16.4%	21.6%	14.8%	11.2%	**68.9%**	**67.2%**
13	**20.9%**	**33.1%**	**34.9%**	**32.4%**	44.2%	34.4%
14	**70.0%**	**72.6%**	10.0%	5.4%	20.0%	22.0%
15	26.7%	34.2%	0.0%	0.0%	**73.3%**	**65.8%**
16	**91.7%**	**96.7%**	0.0%	0.0%	8.3%	3.3%
17	45.3%	44.1%	5.7%	2.9%	**49.1%**	**53.0%**
18	**53.6%**	**56.3%**	0.0%	0.0%	46.4%	43.7%
19	0.0%	0.0%	0.0%	0.0%	**100.0%**	**100.0%**
20	0.0%	0.0%	5.0%	3.1%	**95.0%**	**96.9%**
21	27.0%	21.5%	13.5%	16.5%	**59.5%**	**62.0%**
Chapter average	47.1%	53.1%	10.7%	8.3%	42.2%	38.6%

Bold type: dominant category in chapter

b. Failure analysis

We reviewed all 640 ICD-10-CM codes with partial representation and found three types of reasons for not achieving full representation. (Table 3)

i. Missing information in postcoordination

We identified three kinds of limitation in postcoordination:

- **Postcoordination not allowed**. For example, ICD-10-CM code *H93.13 Tinnitus, bilateral* was recoded as ICD-11 code *MC41Tinnitus* which did not allow postcoordination. Full representation could have been achieved by adding the extension code *XK9J Bilateral*.
- **Addition of existing extension code not allowed**. For example, ICD-10-CM code *M25.552 Pain in left hip* was recoded by postcoordination as *ME82 Pain in joint & XA4XS4 Hip joint*. Full representation could have been achieved if further addition of the extension code *XK8G Left* was allowed.
- **Missing extension code**. Most ICD-10-CM codes for injury and poisoning (chapter 19) included episode of care information (e.g., *S00.31XA Abrasion of nose,* ***initial encounter***), which could not be captured in ICD-11 because there was no extension code for episode of care. Another example was adverse reactions caused by drugs, which had different codes in ICD-10-CM according to the mode of exposure ("adverse effect" if the drug was properly administered, "poisoning" for improper use and "underdosing" if taking less than required). In ICD-11, there was no such distinction and only a general code *NE60 Harmful effects of drugs, medicaments or biological substances, not elsewhere classified* was available. Mode of exposure could be captured by adding three extension codes. Similarly, capturing trimester of pregnancy would require three new

extension codes. Table 3 lists the type of missing information in postcoordination. Note that one code could be associated with more than one type of missing information.

ii. Residual categories

Both ICD-10-CM and ICD-11 have residual categories that usually have "other" or "not elsewhere classified" in their names. These codes are "catch-all" codes to ensure coding of every possible case. The meaning of residual codes can change depending on the neighboring codes, especially the siblings. Therefore, unless all surrounding codes are identical, residual codes from ICD-10-CM and ICD-11 cannot be assumed to be equivalent. Consider for example the ICD-10-CM code *H26.8 Other specified cataract* and ICD-11 code *9B10.2Y Other specified cataracts*. *H26.8* has a sibling *H26.3 Drug-induced cataract* while *9B10.2Y* does not. This means that drug-induced cataract will be included in *9B10.2Y* but not *H26.8,* so *9B10.2Y* is only a partial representation of *H26.8,* despite their exactly matching names.

iii. ICD-11 code more specific than ICD-10-CM code

As mentioned above, we normally used an ICD-11 code that was equivalent to or broader than the ICD-10-CM code, and then refined that code with postcoordination if necessary. In some cases, the ICD-11 coding guidance pointed to a code more specific than the ICD-10-CM code. For example, the ICD-10-CM code *M62.82 Rhabdomyolysis* was recoded to the narrower code *FB32.20 Idiopathic rhabdomyolysis* because the ICD-11 index term "rhabdomyolysis NOS" pointed to this code. We considered these cases partial representation, because idiopathic rhabdomyolysis is more specific than rhabdomyolysis. Of note, postcoordination is not applicable here because postcoordination can only refine the meaning of a broad code but cannot

make a narrow code broader (e.g., postcoordination cannot remove the "idiopathic" characterization from *FB32.20*).

Table 3. Analysis of failure of full representation even with postcoordination

Reason for failure of full representation		**Number of ICD-10-CM codes (%)**
A. Missing information in postcoordination	Episode of care	375 (39.8%)
	Laterality	53 (5.6%)
	Mode of exposure	35 (3.7%)
	Trimester of pregnancy	16 (1.7%)
	Other missing information - anatomy - devices - injury dimension - etiology - substances - severity - temporality - external cause - histopathology - capacity context - others Total	 45 (4.8%) 25 (2.7%) 25 (2.7%) 16 (1.7%) 11 (1.2%) 10 (1.1%) 5 (0.5%) 4 (0.4%) 3 (0.3%) 1 (0.1%) 100 (10.6%) 245 (26.0%)
B. Residual categories		131 (13.9%)
C. ICD-11 more specific		13 (1.4%)

As shown in Table 3, four types of missing information (episode of care, laterality, mode of exposure and trimester of pregnancy) accounted for a large proportion of codes that could not be fully postcoordinated. This would be relatively easy to address, by adding nine extension codes (three episodes of care, three trimesters and three modes of exposure) and allowing the addition of existing laterality modifiers to applicable anatomical entities. If these "easy fixes" were

implemented, the number of ICD-10-CM codes that could be fully-postcoordinated would increase from 81 (8.6%) to 332 (35.2%).

c. Coding guidance review

i. Definitions, inclusions and exclusions

We found no conflicts between the ICD-11 definitions and ICD-10-CM inclusion or exclusion terms. We found 10 cases of conflict among the inclusion and exclusion terms between ICD-10-CM and ICD-11. (Table 4, left half) In one case, there was actual conflict which required changing the target ICD-11 code. In this case, the ICD-10-CM code *O99.820 Streptococcus B carrier state complicating pregnancy* was originally recoded to *JA65.Y Maternal care for other specified conditions predominantly related to pregnancy*. One of the ancestors of the ICD-11 code had an exclusion "Maternal infectious diseases classifiable elsewhere but complicating pregnancy, childbirth or the puerperium (JB63)" which indicated that the code *JB63* should be used instead of *JA65.Y*. The other conflicts were potential ones which only occurred in some specific situations, and the chosen ICD-11 code was generally correct. These potential conflicts belonged to three types:

- *Partial overlap* – an inclusion in one classification occurred as an exclusion in the other classification that pointed to another **broad** code different from the original code. For example, the ICD-10-CM code *A41.9 Sepsis, unspecified organism* was recoded as the ICD-11 code *1G40 Sepsis without septic shock*. "Septicemia" was an inclusion for *A41.9* but an exclusion for *1G40*. In ICD-11, "septicemia" pointed to *MA15 Microbiological findings in blood, blood-forming organs, or the immune system*. The ICD-11 code *1G40*

was correct in the broader context of sepsis. However, in the special case of septicemia one should use *MA15*.

- *Granularity difference* – an inclusion in one classification occurred as an exclusion in the other classification that pointed to a **specific** code different from the original code. For example, the ICD-10-CM code *K59.00 Constipation, unspecified* was recoded as ICD-11 code *ME05.0 Constipation*. "Fecal impaction" was an exclusion for *K59.00* but an inclusion for *ME05.0.* In ICD-10-CM, "fecal impaction" pointed to the more specific code *K56.41 Fecal impaction*. In this case, ICD-10-CM was finer-grained and had distinct codes for specific causes of constipation, but the recoding was correct at the broader level.

- *Default assumption* - an inclusion in one classification occurred as an exclusion in the other classification, but the chosen code was correct according to certain default assumptions. For example, the ICD-10-CM code *R73.09 Other abnormal glucose* was recoded as *5A40.Z Intermediate hyperglycaemia, unspecified* in ICD-11. In ICD-11, *5A40.Z* was defined as "a metabolic disorder characterized by glucose levels too high to be considered normal, though not high enough to meet the criteria for diabetes", i.e., a kind of prediabetes. In ICD-10-CM, *R73.09* had an inclusion "abnormal glucose NOS" and in ICD-11, *5A40.Z* had an exclusion "elevated blood glucose level". Even though the two inclusion/exclusion terms were not exactly the same, one subsumed the other and they were considered in conflict. In ICD-11, the exclusion term "elevated blood glucose level" pointed to *MA18.0 Elevated blood glucose level* (a finding, not a metabolic

disorder). However, the original recoding was considered correct because, in the ICD-11 index, "abnormal glucose" pointed to *5A40.Z*, indicating that unspecified abnormal glucose was coded in ICD-11 as a metabolic disorder by default.

Table 4. Conflicts discovered by review of inclusion, exclusion and index terms

Type of conflict	**Inclusion and exclusion terms**		**Index terms**	
	Number of conflicts	Unique ICD-10-CM codes affected	Number of conflicts	Unique ICD-10-CM codes affected
Actual conflict – target ICD-11 code changed	1	1	8	3
Potential conflict				
1. partial overlap	6	6	109	41
2. granularity difference	2	2	119	54
3. default assumption	1	1	20	12
Total	10	10	266	93

iii. Index terms

We found 266 cases of conflict, eight of which were real conflicts requiring the change of three target ICD-11 codes. (Table 4, right half) For example, the ICD-10-CM code *B19.20 Unspecified viral hepatitis C without hepatic coma* was originally recoded as *1E5Z Viral hepatitis, unspecified*. "Hepatitis C" was indexed to *B19.20* in ICD-10-CM but to *1E51.1 Chronic hepatitis C* in ICD-11, indicating that unspecified hepatitis C was coded as chronic hepatitis C in ICD-11,

so we changed the target code to *1E51.1.* The other cases were potential conflicts. Some examples:

- *Partial overlap* - ICD-10-CM code *Q25.0 Patent ductus arteriosus* was recoded as *LA8B.4 Patent arterial duct*. "Patent ductus arteriosus aneurysm" was indexed to *Q25.0* in ICD-10-CM but to *LA8B.Y Other specified congenital anomaly of great arteries including arterial duct* in ICD-11.

- *Granularity difference* – ICD-10-CM code *L60.3 Nail dystrophy* was recoded as *EE10.5 Nail dystrophy, not otherwise specified*. "Spoon nail" was indexed to *L60.3* in ICD-10-CM but to *EE10.0 Abnormality of nail shape* in ICD-11. In this case, ICD-11 had finer-grained codes for different kinds of nail abnormality compared to ICD-10-CM.

- *Default assumption* – ICD-10-CM code *O03.9 Complete or unspecified spontaneous abortion without complication* was recoded to *JA00.09 Spontaneous abortion, complete or unspecified, without complication*. While "abortion" was indexed to *O03.9*, indicating that ICD-10-CM assumed spontaneous abortion by default, in ICD-11, "abortion" was indexed to a more general code *JA00.2 Unspecified abortion*.

DISCUSSION

1. Coverage of ICD-10-CM by ICD-11

Compared to our previous study, [10] this is a more comprehensive appraisal of the feasibility of replacing ICD-10-CM with ICD-11. Of the 943 codes that we studied, representing the most

frequently used codes in each chapter, 23.5% could be fully represented without postcoordination, and a further 8.6% with postcoordination. Analysis of the partially represented codes revealed that a few types of missing information accounted for a large number of cases. These are the “low-hanging fruits” for improving the alignment between ICD-10-CM and ICD-11. With the implementation of some “easy fixes” (adding nine extension codes and allowing laterality qualification of anatomic entities), the number of codes that can be fully postcoordinated would increase from 8.6% to 35.2%, bringing the proportion of full representation from 32.1% to 58.7%. Further improvement in coverage can be achieved by adding more extension codes, though with diminishing returns. According to our analysis, adding extension codes in the sub-branches of anatomy, devices and injury dimensions among the ICD-11 extension codes will have the largest impact.

Both ICD-10-CM and ICD-11 provide coding guidance in terms of inclusion and exclusion terms, as well as an index. These are important references in defining the meaning and the boundaries of a code. Even when an ICD-10-CM code has exactly the same description as an ICD-11 code, nuances in meaning can still exist as indicated by the coding guidance. In this study, we did find conflicts in the coding guidance analysis. This resulted in the need to change the target ICD-11 code in a small number of cases. In about 10% of the codes, we detected potential conflicts, which, even though they did not invalidate the matching with the ICD-11 code, would require a change of code in certain situations (e.g., when a specific condition encompassed by a code in one classification is coded differently in the other.) This shows that the accompanying coding guidance should be taken into consideration when assessing the alignment between ICD-10-CM and ICD-11.

2. ICD-11 as a replacement for ICD-10-CM

One important consideration in changing from one coding system to another is the amount of disruption in coding. Based on the 2016 General Equivalence Maps (GEMs) published by CMS immediately after the transition to ICD-10-CM, of 14,567 ICD-9-CM codes, only 3,533 (24.3%) had an exact match in ICD-10-CM. [16] This is very close to the 23.5% full representation of ICD-10-CM codes by ICD-11 codes without postcoordination found in this study. With postcoordination and some minor enhancements, full representation would further increase to 58.7%. Based on this, moving from ICD-10-CM to ICD-11 appears much less disruptive than moving from ICD-9-CM to ICD-10-CM. Therefore, before embarking on the development of ICD-11-CM, serious consideration should be given to using ICD-11 for morbidity coding, possibly enhanced with postcoordination. One caveat is that postcoordination is a brand-new feature in ICD-11. It has never been used in ICD coding and will have impact on tooling, coder education and coding variability. In our study, the inter-coder agreement for postcoordination is comparable to the selection of the main codes.

Using ICD-11 for morbidity coding in the US would avoid the cost of maintaining a national extension and the potential divergence of meaning from the international core. There are also benefits of an up-to-date medical nomenclature that reflects state-of-the-art biomedical knowledge. Overall, content coverage is only one of the factors to consider in deciding whether ICD-11 can replace ICD-10-CM. Other factors such as cost and benefit analysis, resource impact and burden of implementing ICD-11 for morbidity coding are beyond the scope of this study.

3. Limitations and future work

We recognize the following limitations in this investigation. Our list of frequently used codes was derived from Medicare claims data and three hospitals. Even though the total number of patients covered was substantial, it may not be representative of all healthcare settings. Our index terms review was based on conflicts detected by normalized lexical matching because of the impracticality of reviewing all index terms. This had resulted in a small number of false positives (e.g., “baby blues” and “blue baby” were normalized to the same term), which we detected in our manual review. However, there could be false negatives (missed by lexical matching) which we did not know about. The new foundation component of ICD-11 can confer additional benefits such as easier integration with SNOMED CT and other standard terminologies stipulated in the Promoting Interoperability and related initiatives. [17-20] In the future, we will investigate how the foundation component in ICD-11 can help to align ICD-11 with other terminology standards.

CONCLUSION

Based on the analysis of the 943 most frequently used ICD-10-CM codes covering 60% of usage, 222 (23.5%) could be fully represented without postcoordination, 81 codes (8.6%) could be fully represented with postcoordination, the remaining 640 codes (67.9%) could only achieve partial representation. With minor changes to ICD-11, the proportion that can be fully represented would increase to 58.7%. Analysis of the inclusions, exclusions and index revealed potential conflicts in 10% of codes, but few actual conflicts affecting recoding.

Acknowledgements

The authors would like to thank James Campbell and Ellen Kerns for providing data from the University of Nebraska Medical Center. We thank the National Committee on Vital and Health Statistics for suggesting and promoting research into the feasibility of using ICD-11 for morbidity coding without a US Clinical Modification.

Funding Statement

This research was supported in part by the Intramural Research Program of the NIH, National Library of Medicine.

Competing Interests Statement

The authors do not have competing interests. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Library of Medicine, National Institutes of Health, National Center for Health Statistics or the Centers for Disease Control and Prevention.

Contributorship Statement

KWF, DP and OB conceived and designed the study. JX and SM performed the recoding of ICD-10-CM codes and reviewed the coding guidance. KWF performed the data analysis. KWF drafted the manuscript and all authors contributed substantially to its revision.

References

1. *History of the development of the ICD, World Health Organization,* *http://www.who.int/classifications/icd/en/HistoryOfICD.pdf*.
2. Gersenovic, M., *The ICD family of classifications.* Methods Inf Med, 1995. **34**(1-2): p. 172-5.
3. *World Health Assembly Update, 25 May 2019: International Statistical Classification of Diseases and Related Health Problems (ICD-11)*. 2019; Available from: https://www.who.int/news-room/detail/25-05-2019-world-health-assembly-update.
4. Pocai, B., *The ICD-11 has been adopted by the World Health Assembly.* World Psychiatry, 2019. **18**(3): p. 371-372.
5. Editorial, *ICD-11: a brave attempt at classifying a new world.* Lancet, 2018. **391**(10139): p. 2476.
6. Centers for Disease Control and Prevention. *Scientific Data Documentation: International Classification of Diseases - 9 - CM, (1979)*. Available from: https://wonder.cdc.gov/wonder/sci_data/codes/icd9/type_txt/icd9cm.asp.
7. Cimino, J.J., *Desiderata for controlled medical vocabularies in the twenty-first century.* Methods Inf Med, 1998. **37**(4-5): p. 394-403.
8. *ICD-11 Home Page*. Available from: https://icd.who.int/en/.
9. *ICD-11 Reference Guide*. Available from: https://icd.who.int/icd11refguide/en/index.html.
10. Fung, K.W., J. Xu, and O. Bodenreider, *The new International Classification of Diseases 11th edition: a comparative analysis with ICD-10 and ICD-10-CM.* J Am Med Inform Assoc, 2020. **27**(5): p. 738-746.
11. National Committee on Vital and Health Statistics. *Recommendation Letter to Department of Health and Human Services: Preparing for Adoption of ICD-11 as a Mandated U.S. Health Data Standard*. 2019; Available from: https://ncvhs.hhs.gov/wp-content/uploads/2019/12/Recommendation-Letter-Preparing-for-Adoption-of-ICD-11-as-a-Mandated-US-Health-Data-Standard-final.pdf.
12. Research Data Assistance Center (ResDAC). *CMS Virtual Research Data Center (VRDC)*. Available from: https://www.resdac.org/cms-virtual-research-data-center-vrdc.
13. *ICD-11 for Mortality and Morbidity Statistics Browser*. Available from: https://icd.who.int/browse11/l-m/en.
14. McCray, A.T., S. Srinivasan, and A.C. Browne, *Lexical methods for managing variation in biomedical terminologies.* Proc Annu Symp Comput Appl Med Care, 1994: p. 235-9.
15. *UMLS Reference Manual: SPECIALIST Lexicon and Lexical Tools*. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9680/.
16. Centers for Medicare and Medicaid Services, C. *ICD-10-CM and ICD-10 PCS and GEMs Archive*. Available from: https://www.cms.gov/Medicare/Coding/ICD10/Archive-ICD-10-CM-ICD-10-PCS-GEMs.
17. Mamou, M., et al., *Representing ICD-11 JLMMS Using IHTSDO Representation Formalisms.* Stud Health Technol Inform, 2016. **228**: p. 431-5.
18. Rodrigues, J.M., et al., *Semantic Alignment between ICD-11 and SNOMED CT.* Stud Health Technol Inform, 2015. **216**: p. 790-4.
19. Schulz, S., et al., *What's in a class? Lessons learnt from the ICD - SNOMED CT harmonisation.* Stud Health Technol Inform, 2014. **205**: p. 1038-42.
20. *Promoting Interoperability, Centers for Medicare & Medicaid Services*. Available from: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms/.

Figure 1. Agreement of ICD-11 coding between the two terminologists.